Disparities in Defining Disparities: Statistical Conceptual Frameworks

STATISTICS IN MEDICINE
Statist. Med. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3283

Disparities in defining disparities: Statistical conceptual frameworks

Naihua Duan[1, 2, 3, *, †], Xiao-Li Meng[4], Julia Y. Lin[5,6], Chih-nan Chen[6] and Margarita Alegria[5,6]

[1]*Department of Psychiatry, Columbia University, New York, NY 10032, U.S.A.*
[2]*New York State Psychiatric Institute, New York, NY 10032, U.S.A.*
[3]*Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.*
[4]*Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.*
[5]*Department of Psychiatry, Harvard Medical School, Boston, MA 02215, U.S.A.*
[6]*Center for Multicultural Mental Health Research, Somerville, MA 02143, U.S.A.*

SUMMARY

Motivated by the need to meaningfully implement the Institute of Medicine's (IOM's) definition of health care disparity, this paper proposes statistical frameworks that lay out explicitly the needed causal assumptions for defining disparity measures. Our key emphasis is that a scientifically defensible disparity measure must take into account the direction of the causal relationship between *allowable covariates* that are not considered to be contributors to disparity and *non-allowable covariates* that are considered to be contributors to disparity, to avoid flawed disparity measures based on implausible populations that are not relevant for clinical or policy decisions. However, these causal relationships are usually unknown and undetectable from observed data. Consequently, we must make strong causal assumptions in order to proceed. Two frameworks are proposed in this paper, one is the *conditional disparity* framework under the assumption that allowable covariates impact non-allowable covariates but not *vice versa*. The other is the *marginal disparity* framework under the assumption that non-allowable covariates impact allowable ones but not *vice versa*. We establish theoretical conditions under which the two disparity measures are the same and present a theoretical example showing that the difference between the two disparity measures can be arbitrarily large. Using data from the Collaborative Psychiatric Epidemiology Survey, we also provide an example where the conditional disparity is misled by Simpson's paradox, whereas the marginal disparity approach handles it correctly. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: counterfactual populations; disparities; potential outcomes; weighting; mental health; Simpson's paradox

*Correspondence to: Naihua Duan, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 636, New York, NY 10032, U.S.A.
†E-mail: naihua.duan@columbia.edu

Contract/grant sponsor: NIH; contract/grant numbers: P50-MHO73469-03, U01-MH06220-06A2

Received 17 February 2008
Accepted 19 February 2008

1. CAUSALITY AND DISPARITY MEASURES

1.1. The causal implication of the IOM definition

The Institute of Medicine (IOM) [1] defines health care disparities as 'racial or ethnic differences in the quality of health care that are not due to access-related factors or clinical needs, preference, and appropriateness of intervention.' This definition represents an important advance in disparity research, because it explicitly recognizes the role of causality in the determination of disparities through its reference to the causal expression 'not due to'. However, it leaves open the interpretation of the causal model underlying this causal statement. In this paper we identify several causal models under which the IOM definition can be implemented meaningfully and propose the corresponding frameworks for defining and comparing statistically justifiable disparity measures following these models. Our work can be viewed as a statistically oriented conceptualization of research in this area (e.g. [2–6]). Although our work was directly motivated by the IOM definition, the proposed general frameworks are equally applicable to other areas, such as in legal settings (e.g. [7–9]).

The statistical frameworks proposed in this paper assume that the covariates of interest have been classified into *allowable* and *non-allowable* categories. Allowable covariates are considered to be justifiable to cause a difference and hence should be adjusted before measuring disparity. The remaining covariates are classified as non-allowable.

It is important to note that the classification of allowable and non-allowable covariates can, and should, vary from study to study, depending on the particular purpose of the study. For example, IOM's classification of access-related factors as allowable is appropriate for studying disparity at the level of patient-clinician encounter, with the focus being the treatment delivered during the encounter, controlled for all historical factors that occurred prior to the encounter. However, when studying health care disparity at the level of service systems, it would be more appropriate to classify access-related factors as non-allowable, thus holding the service systems accountable for failure to engage disadvantaged patients into care. The statistical frameworks we establish in this paper apply to any of such classifications.

As a specific example for illustration, suppose that covariates that might be predictive of health care are classified as follows:

- Clinical needs and preference are considered allowable. Differences in health care due to these covariates are *not* considered to be part of health care disparity.
- All other covariates, such as knowledge about health, state of residency, insurance coverage, and education (to name a few), are considered non-allowable. Differences in health care due to these covariates are considered to be health care disparity.

Given such a classification, our goal then is to measure the disparity that is '*not due to*' the allowable covariates.

A seemingly obvious, and hence very common, approach is to substitute the levels of allowable covariates of, for example, Afro-Caribbean with those of their non-Latino white counterparts, while leaving the levels of non-allowable covariates unchanged. This procedure is often used in the analysis of covariance models that adjust for allowable covariates across racial/ethnic groups. However, this approach is sensible in general terms only if the allowable covariates are statistically independent of the non-allowable covariates, a condition that is unlikely to hold in practice. Without this independence condition, this direct substitution may lead to an implausible population, such as a hypothetical population with high level of income (as a non-allowable covariate that remains

DOI: 10.1002/sim

unchanged) and a high level of chronic diseases (as an allowable covariate that was substituted with the levels from the reference population). As a result, the disparity estimates obtained from this procedure may not be relevant for clinical, policy, or other purposes, because they are based on a postulated population that cannot be realized by policy changes or disparities interventions.

Not accounting properly for the causal relationships between allowable and non-allowable covariates is especially problematic when the two sets of covariates are highly correlated in the observed data, and both sets of variables are included in our outcome model. In such cases, the allowable covariates might appear to be very weak for predicting the outcome in the fitted model due to the well-known 'collinearity' phenomenon. Consequently, replacing a minority group's allowable covariates by their counterparts in the non-Latino white group in the fitted model may produce only trivial adjustment, even if in reality a substantial part of the observed racial/ethnic difference is indeed due to the difference in the allowable covariates. This could be because of their direct impact either on the outcome (which would not be detected by the fitted regression model because of the strong collinearity) or on the non-allowable covariates or on both. The frameworks proposed in this paper can help to substantially reduce such serious mis-estimation of disparity because they explicitly take into account the causal relationship between the allowable and the non-allowable covariates. For example, our approaches permit an adjustment in allowable covariates to cause substantial adjustment in the non-allowable ones, which in turn may lead to substantial adjustment in the predicted outcome, even if the allowable predictor appears to be very weak in the fitted model for predicting the outcome.

1.2. Explicating the underlying causal assumptions

In order to measure disparity meaningfully, such as to implement the IOM definition for health care disparity, one must be explicit about the underlying causal assumptions that are imbedded in any disparity measure. The fact that the exact causal mechanisms may not be known or may not even be knowable is not a reason to 'sweep everything under the rug.' On the contrary, this is precisely the reason for us to be explicit about our assumptions, so that it is possible for policy makers and subsequent researchers to correctly interpret the disparity measures/estimates we obtain as well as to determine the directions for correction or improvement when newer information becomes available for the underlying causal relationships.

The key reason why we need to make causal assumptions is that once an action is forced upon a particular variable (e.g. by changing a minority group's distribution of clinical needs to match those of the non-Latino white population), it will have a ripple effect—in real life—on other variables (e.g. income level) that are impacted by the one adjusted. However, this ripple effect is not estimable without carrying out the actual (social) experiment, because the observed relationships in a natural population may or may not be preserved after an intervention. As an illustrative example, in a natural population, a person's left-eye visual acuity (VA) may be highly correlated with the person's right-eye VA. However, this correlation will be destroyed or at least reduced if we perform a vision correction laser surgery on the right eye only. The two VAs will become independent shortly after the surgery but may become correlated again over time, though the cross-sectional data from a natural population would tell us little about how large this correlation could be or whether it would ever reach the same level as in the natural population.

Therefore, in order to measure the disparity 'not due to' the allowable covariates, we must postulate causal directions as well as how relationships among relevant variables are preserved or altered with the change from a natural population to a hypothetical one. There are two extreme types

DOI: 10.1002/sim

of unidirectional causal relationships: (A) allowable covariates impact non-allowable covariates but not *vice versa* and (B) non-allowable covariates impact allowable covariates but not *vice versa*. The more realistic relationships are likely to be either (C) allowable covariates and non-allowable covariates are inter-related and reciprocally impact each other, or (D), which is (C) plus the possibility that both allowable and non-allowable covariates are also impacted by the outcome itself (over time).

Although (C) and (D) are most dynamic and realistic, they do not permit useful modeling without further specifications on how the variables involved impact each other. As these specifications are content dependent and can be extremely difficult to postulate, we will pursue them in future work. In this paper, we lay out the statistical frameworks for the simpler causal relationships (A) and (B). These two frameworks serve as building blocks for more complex causal specifications and at the same time provide plausible specifications that might yield useful bounds on the true disparity when more complicated causal relationships are present. In some applications, such as the one presented in Section 3.2, such simplistic causal assumptions are actually reasonable, leading to sensible practical solutions.

2. STATISTICAL FRAMEWORKS

2.1. Linking natural and hypothetical joint distributions

Let $\mathbf{X}_{\mathrm{N}}$ denote non-allowable covariates such as knowledge about health, and let $\mathbf{X}_{\mathrm{A}}$ denote allowable covariates such as clinical needs. Let Y denote the outcome of interest, such as log of the health care expenditure. To measure the disparity, we need to adjust the levels of allowable covariates ($\mathbf{X}_{\mathrm{A}}$) but not the levels of non-allowable covariates ($\mathbf{X}_{\mathrm{N}}$). Note here that all variables are measured for each individual i, but we suppress the subscript i throughout the text to simplify the notation. To describe the distribution of these variables, we use the common generic notation $P(\ )$, e.g. $P(\mathbf{X}_{\mathrm{N}})$. Whenever needed, we will use subscript 1 to denote the reference group (e.g. the non-Latino whites) and 2 the group of interest (e.g. a minority group), for example, $P_1(\mathbf{X}_{\mathrm{N}})$ and $P_2(\mathbf{X}_{\mathrm{N}})$.

The goal of our modeling is to estimate the potential outcome that would be manifested if the group of interest had the same levels of allowable covariates as the reference group. The first step in setting up our proposed frameworks is to explicitly consider the joint distribution of $(Y, \mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{N}})$ and recognize that there are two joint distributions of interest: one for the natural population and the other for the adjusted hypothetical population. We use the superscript (H) to denote different populations, e.g. $P_2^{(H)}(\ )$, where (H) can refer to either an adjustment rule for a hypothetical population (e.g. $P_2^{(A)}(\ )$ for adjustment rule (A)) or a natural (or non-adjusted) population (e.g. $P_2^{(N)}(\ )$). For any (H), we always have the following decomposition:

$$P_2^{(H)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) = P_2^{(H)}(Y|\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})\, P_2^{(H)}(\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) \tag{1}$$

The importance of recognizing the dependence on H here is that only the natural population, $P^{(N)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})$, can be estimated from the data. Therefore, in order to calculate disparities under a hypothetical population, we need to make strong assumptions to link the hypothetical population, such as $P_2^{(A)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})$, to the natural population $P_2^{(N)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})$. Our first assumption, which appears to be taken for granted in much of the existing literature, is that the 'forced action' of the

DOI: 10.1002/sim

adjustment has no impact on the conditional distribution of Y given $(\mathbf{X}_N, \mathbf{X}_A)$. That is, for any adjustment rule (A), we assume

$$P_2^{(A)}(Y|\mathbf{X}_N, \mathbf{X}_A) = P_2^{(N)}(Y|\mathbf{X}_N, \mathbf{X}_A) \tag{2}$$

We will refer to (2) as the 'predictively nature preserving' (PNP) assumption, meaning that the predictive nature of $\{\mathbf{X}_N, \mathbf{X}_A\}$ on Y is preserved despite the 'forced action' on $\mathbf{X}_A$.

One can easily consider a scenario under which the PNP assumption is false, but without such an assumption, the estimation of the disparity is essentially impossible. For example, in our hypothetical eye vision example, two people may have identical VAs for both eyes (e.g. both are 20/20 in the right eye but 20/40 in the left eye), but they can have quite different probabilities of having automobile accidents if one of them was born with such vision, but the other achieved it via laser surgery to his right eye. Clearly, if this occurs, then it is impossible to estimate—using only the data collected from the natural population—the accident rate for the group of people with vision corrections done to their right eyes only.

To carry the decomposition (1) further, we can decompose the component $P_2^{(H)}(\mathbf{X}_N, \mathbf{X}_A)$ in (1) into one conditional distribution and one marginal distribution. This time, there are two possibilities:

$$P_2^{(H)}(\mathbf{X}_N, \mathbf{X}_A) = P_2^{(H)}(\mathbf{X}_N|\mathbf{X}_A)\, P_2^{(H)}(\mathbf{X}_A) \tag{3}$$

and

$$P_2^{(H)}(\mathbf{X}_N, \mathbf{X}_A) = P_2^{(H)}(\mathbf{X}_A|\mathbf{X}_N)\, P_2^{(H)}(\mathbf{X}_N) \tag{4}$$

The first decomposition is the basis for our *conditional* framework, which assumes that non-allowable covariates $\mathbf{X}_N$ are causally dependent on allowable covariates $\mathbf{X}_A$ but not *vice versa*. The second decomposition is suitable for the *marginal* causal framework, which assumes that the allowable covariates $\mathbf{X}_A$ are causally dependent on the non-allowable covariates $\mathbf{X}_N$ but not *vice versa*. Below we show how we can create different counterfactual populations, a standard practice in causal inferences (e.g. see [10]), using these decompositions.

2.2. *Conditional disparity*

Under the conditional framework, we adjust the marginal distribution of the allowable covariates $\mathbf{X}_A$ from the natural population (such as Latinos) to the corresponding marginal distribution of the reference group (such as non-Latino whites), while preserving the conditional distribution for non-allowable covariates $\mathbf{X}_N$ given allowable covariates $\mathbf{X}_A$ as in the natural population. Specifically, the hypothetical joint distribution is obtained by replacing the marginal distribution of $\mathbf{X}_A$ in the natural population

$$P_2^{(N)}(Y, \mathbf{X}_N, \mathbf{X}_A) = P_2^{(N)}(Y|\mathbf{X}_N, \mathbf{X}_A)\, P_2^{(N)}(\mathbf{X}_N|\mathbf{X}_A)\, P_2^{(N)}(\mathbf{X}_A) \tag{5}$$

by that of the reference population (e.g. non-Latino whites), thereby creating the following hypothetical population distribution:

$$P_2^{(C)}(Y, \mathbf{X}_N, \mathbf{X}_A) = P_2^{(N)}(Y|\mathbf{X}_N, \mathbf{X}_A)\, P_2^{(N)}(\mathbf{X}_N|\mathbf{X}_A)\, P_1^{(N)}(\mathbf{X}_A) \tag{6}$$

Although $P_1^{(N)}(\mathbf{X}_A)$ is taken from the natural population of the reference group, its insertion into (5) leads to a hypothetical population that retains the natural conditional distributions $P_2^{(N)}(Y|\mathbf{X}_N, \mathbf{X}_A)$

DOI: 10.1002/sim

and $P_2^{(N)}(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A})$, with the component $P_2^{(N)}(\mathbf{X}_\mathrm{A})$ 'mutated' into $P_1^{(N)}(\mathbf{X}_\mathrm{A})$. We denote this adjustment rule under the *conditional* disparity framework as adjustment (C).

In order for (6) to be a meaningful hypothetical population, our assumptions are as follows: (i) the PNP assumption holds and (ii) the adjustment action has no impact on the conditional distribution of $\mathbf{X}_\mathrm{N}$ given $\mathbf{X}_\mathrm{A}$ either; that is,

$$P_2^{(C)}(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A}) = P_2^{(N)}(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A}) \tag{7}$$

which is plausible when the causal direction is from $\mathbf{X}_\mathrm{A}$ to $\mathbf{X}_\mathrm{N}$ but not *vice versa*. We will refer to (7) as the 'conditionally nature preserving' (CNP) assumption, meaning that the natural conditional distribution $P_2(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A})$ is preserved after the adjustment on $\mathbf{X}_\mathrm{A}$.

The ratio between the adjusted joint density (6) and the natural joint density (5) is simply the ratio of the marginal densities:

$$R_\mathrm{C}(\mathbf{X}_\mathrm{A}) = \frac{P_1^{(N)}(\mathbf{X}_\mathrm{A})}{P_2^{(N)}(\mathbf{X}_\mathrm{A})} \tag{8}$$

Following the principle of importance weighting, the expected outcome under the hypothetical population (6) can be expressed as the following *weighted* expectation of Y under the natural population (5), with the importance weight $R_\mathrm{C}(\mathbf{X}_\mathrm{A})$:

$$E_2^{(C)}[Y] = E_2^{(N)}[Y R_\mathrm{C}(\mathbf{X}_\mathrm{A})] \tag{9}$$

where $E_2^{(C)}$ denotes the expectation with respect to the hypothetical population in (6) and $E_2^{(N)}$ denotes the expectation with respect to the natural population in (5).

Expression (9) gives us a practical way to estimate $E_2^{(C)}[Y]$ because its right-hand side involves only expectations with respect to the natural population (5), therefore it can be estimated from the sample data. As this paper focuses on setting up conceptual frameworks, the detailed estimation procedures, particularly for estimating $R_\mathrm{C}(\mathbf{X}_\mathrm{A})$, will be presented in a subsequent paper.

Intuitively, the adjustment under our conditional framework amounts to weighting the level of health care (Y) among minorities by the density ratio $R_\mathrm{C}(\mathbf{X}_\mathrm{A})$. Minorities with higher density ratio R_C get weighted up because a value of $R_\mathrm{C}(\mathbf{X}_\mathrm{A})>1$ tells us that there are more non-Latino whites with the levels of $\mathbf{X}_\mathrm{A}$ than minorities with the same levels of $\mathbf{X}_\mathrm{A}$. The corresponding disparity is then measured as the difference between the expected value of Y for the adjusted (hypothetical) population (6) and that of the reference population:

$$D_\mathrm{C} = E_2^{(C)}[Y] - E_1^{(N)}[Y] \tag{10}$$

We term D_C of (10) as *conditional disparity* because the main source of disparity is in the difference in the *conditional distributions* $P_2^{(N)}(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A})$ and $P_1^{(N)}(\mathbf{X}_\mathrm{N}|\mathbf{X}_\mathrm{A})$. The difference in $P_2^{(N)}(Y|\mathbf{X}_\mathrm{N},\mathbf{X}_\mathrm{A})$ and $P_1^{(N)}(Y|\mathbf{X}_\mathrm{N},\mathbf{X}_\mathrm{A})$ may also be of interest in its own right, an issue we shall not pursue here due to the page limitation, but will briefly touch upon in Section 3.3.

Applying expression (9) to definition (10), we have the following expression for conditional disparity that can be estimated using sample data:

$$D_\mathrm{C} = E_2^{(N)}[Y R_\mathrm{C}(\mathbf{X}_\mathrm{A})] - E_1^{(N)}[Y] \tag{11}$$

DOI: 10.1002/sim

Note that this expression for conditional disparity does not involve the non-allowable covariates, X_{N}. This is possible because of the assumption that $\mathbf{X}_{\mathrm{N}}$ is caused by $\mathbf{X}_{\mathrm{A}}$. Under such an assumption, we can greatly simplify the estimation task because (11) bypasses the need to model X_{N}. The theoretical implication of this simplification will be discussed in Section 4.

2.3. Marginal disparity

In contrast to conditional disparity, which equates the two marginal distributions of $\mathbf{X}_{\mathrm{A}}$, the marginal disparity framework replaces the conditional distribution of $\mathbf{X}_{\mathrm{A}}$, conditioning on $\mathbf{X}_{\mathrm{N}}$, of the population of interest (e.g. Latinos) by that of the reference population (e.g. non-Latino whites). Specifically, we replace the conditional distribution $P_2^{(N)}(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})$ in the natural population

$$P_2^{(N)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) = P_2^{(N)}(Y|\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})\, P_2^{(N)}(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})\, P_2^{(N)}(\mathbf{X}_{\mathrm{N}}) \tag{12}$$

by that of the reference population to create the following hypothetical population:

$$P_2^{(M)}(Y, \mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) = P_2^{(N)}(Y|\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})\, P_1^{(N)}(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})\, P_2^{(N)}(\mathbf{X}_{\mathrm{N}}) \tag{13}$$

We denote this adjustment rule under the *marginal* disparity framework as adjustment (M).

Similar to the conditional disparity framework, in order for (13) to be a meaningful hypothetical population, we have assumed that (i) the PNP assumption holds and (ii) the adjustment action has no impact on the marginal distribution of $\mathbf{X}_{\mathrm{N}}$ either; that is,

$$P_2^{(M)}(\mathbf{X}_{\mathrm{N}}) = P_2^{(N)}(\mathbf{X}_{\mathrm{N}}) \tag{14}$$

which is plausible when the causal direction is from $\mathbf{X}_{\mathrm{N}}$ to $\mathbf{X}_{\mathrm{A}}$ but not *vice versa*. We will refer to (14) as the 'marginally nature preserving' (MNP) assumption, meaning that the marginal distribution $P_2(\mathbf{X}_{\mathrm{N}})$ is preserved after the adjustment on $\mathbf{X}_{\mathrm{A}}$.

Similar to (8), the ratio between the joint densities (13) and (12) is given by the ratio between the two conditional densities:

$$R_{\mathrm{M}}(\mathbf{X}_{\mathrm{A}}; \mathbf{X}_{\mathrm{N}}) = \frac{P_1^{(N)}(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})}{P_2^{(N)}(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})} \tag{15}$$

Again, the ratio (15) can be used as the importance weight to express

$$E_2^{(M)}[Y] = E_2^{(N)}[Y\, R_{\mathrm{M}}(\mathbf{X}_{\mathrm{A}}; \mathbf{X}_{\mathrm{N}})] \tag{16}$$

where $E_2^{(M)}$ denotes expectation under the hypothetical population (13) and $E_2^{(N)}$ denotes expectation under the natural population (12). Note that the right-hand side of (16) can be estimated from sample data obtained in the natural population (12).

It is useful to visualize the adjustment under our marginal framework as first stratifying the minority population by the level of the non-allowable covariates (e.g. knowledge of health). We then apply the same weighting scheme as with the conditional disparity approach but now *within each stratum*; therefore the weights there, namely, the ratio of marginal densities $R_{\mathrm{C}}(\mathbf{X}_{\mathrm{A}})$ is now replaced by the ratio of the corresponding conditional densities $R_{\mathrm{M}}(\mathbf{X}_{\mathrm{A}}; \mathbf{X}_{\mathrm{N}})$. Minorities within a particular stratum, as defined by their values of $\mathbf{X}_{\mathrm{N}}$, with higher conditional density ratio R_{M} get weighted up when there are more non-Latino whites with the levels of $\mathbf{X}_{\mathrm{A}}$ than minorities *in the same stratum* as defined by the value of $\mathbf{X}_{\mathrm{N}}$.

DOI: 10.1002/sim

The marginal disparity measure then is defined as the difference between the expected value of Y for the adjusted (hypothetical) population (13) and that of the reference population (12):

$$D_{\mathrm{M}} = E_2^{(M)}[Y] - E_1^{(N)}[Y] \tag{17}$$

We term D_{M} as *marginal disparity* because the main source of the disparity is in the difference in the *marginal distributions* $P_2^{(N)}(\mathbf{X}_{\mathrm{N}})$ and $P_1^{(N)}(\mathbf{X}_{\mathrm{N}})$, in addition to any difference in $P_2^{(N)}(Y|\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})$ and $P_1^{(N)}(Y|\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})$. Again, applying expression (16) to the definition (17), we have the following expression for marginal disparity that can be estimated using sample data:

$$D_{\mathrm{M}} = E_2^{(N)}[Y R_{\mathrm{M}}(\mathbf{X}_{\mathrm{A}}; \mathbf{X}_{\mathrm{N}})] - E_1^{(N)}[Y] \tag{18}$$

The estimation of $R_{\mathrm{M}}(\mathbf{X}_{\mathrm{A}}; \mathbf{X}_{\mathrm{N}})$ is more complicated than estimating $R_{\mathrm{C}}(\mathbf{X}_{\mathrm{A}})$ due to the higher dimensionality. Again, these technical details will be addressed in a subsequent paper.

3. COMPARING CONDITIONAL AND MARGINAL DISPARITIES

With the two frameworks given above, a natural question is when do they give the same disparity estimates, or more profoundly, do they give different values that would matter in practice? The answer to the first part is a clean-cut theoretical result we present below. The answer to the second part is obviously 'it depends' because it depends critically on the nature of the dependence structure between $\mathbf{X}_{\mathrm{A}}$ and $\mathbf{X}_{\mathrm{N}}$, as well as the dependence of Y on $(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{N}})$, in particular applications. We will illustrate this with two examples, one of which shows the difference between getting it right or wrong, and the other gives a class of cases where the difference can be made arbitrarily large. For the remainder of this paper, we suppress the superscript (N) as a notation for the natural population, whenever the context is clear.

3.1. A theoretical result related to local dependence function

The difference between D_{C} and D_{M} can be expressed as

$$\Delta D \equiv D_{\mathrm{C}} - D_{\mathrm{M}} = E_2\left[Y \cdot \left(\frac{P_1(\mathbf{X}_{\mathrm{A}})}{P_2(\mathbf{X}_{\mathrm{A}})} - \frac{P_1(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})}{P_2(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})}\right)\right] \tag{19}$$

The two disparity measures will be identical, $\Delta D = 0$, if

$$\frac{P_1(\mathbf{X}_{\mathrm{A}})}{P_2(\mathbf{X}_{\mathrm{A}})} = \frac{P_1(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})}{P_2(\mathbf{X}_{\mathrm{A}}|\mathbf{X}_{\mathrm{N}})} \tag{20}$$

This condition is equivalent to the condition that

$$G_1(\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) \equiv \frac{P_1(\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})}{P_1(\mathbf{X}_{\mathrm{N}})P_1(\mathbf{X}_{\mathrm{A}})} = \frac{P_2(\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}})}{P_2(\mathbf{X}_{\mathrm{N}})P_2(\mathbf{X}_{\mathrm{A}})} \equiv G_2(\mathbf{X}_{\mathrm{N}}, \mathbf{X}_{\mathrm{A}}) \tag{21}$$

Here the G function can be viewed as a measure of the dependence structure between $\mathbf{X}_{\mathrm{N}}$ and $\mathbf{X}_{\mathrm{A}}$; therefore, condition (21) says that as long as the dependence structure is the same for the two groups (e.g. it remains the same across the two racial/ethnic groups), the two disparity measures would be identical. As a special case, if $\mathbf{X}_{\mathrm{N}}$ and $\mathbf{X}_{\mathrm{A}}$ are independent under *both* populations, then the two measures are the same because both G_1 and G_2 are then identical to 1.

DOI: 10.1002/sim

For continuous variables, the notion that G is a measure of dependence structure can also be examined through the *local dependence function* (LDF), as defined in [11] and studied in [12] and [13],

$$\gamma(\mathbf{X}_N, \mathbf{X}_A) = \frac{\partial^2 \log P(\mathbf{X}_N, \mathbf{X}_A)}{\partial \mathbf{X}_N \partial \mathbf{X}_A} \tag{22}$$

Because

$$\frac{\partial^2 \log G(\mathbf{X}_N, \mathbf{X}_A)}{\partial \mathbf{X}_N \partial \mathbf{X}_A} = \frac{\partial^2 \log P(\mathbf{X}_N, \mathbf{X}_A)}{\partial \mathbf{X}_N \partial \mathbf{X}_A} \tag{23}$$

it is obvious that condition (21) implies that the LDF is independent of the group index, i.e. the LDF does not change with the racial/ethnic group. Note, however, that the reverse is not necessarily true; that is, we can have the LDF invariant to group index, but condition (21) does not hold. In this sense, the measure of dependence defined by the G function is more stringent than that defined by the LDF.

Finally we note that condition (21) is sufficient but not necessary for $\Delta D = 0$. A simple example is that $\Delta D = 0$ when the regression of Y on $\mathbf{X}_A$ and $\mathbf{X}_N$, that is, $E_2[Y|\mathbf{X}_N, \mathbf{X}_A]$, is free of *both* $\mathbf{X}_N$ and $\mathbf{X}_A$ (note that this is a weaker requirement than the independence between Y and $(\mathbf{X}_N, \mathbf{X}_A)$ as only the conditional mean of Y is involved). This, of course, does not happen when $\mathbf{X}_N$ and/or $\mathbf{X}_A$ are useful predictors of Y. However, it reminds us that the difference between D_C and D_M also depends on the relationship between Y and $(\mathbf{X}_N, \mathbf{X}_A)$, and the difference will be small when both $\mathbf{X}_N$ and $\mathbf{X}_A$ are weak predictors.

We emphasize here that the statement we just made is true only when *both* $\mathbf{X}_N$ and $\mathbf{X}_A$ are weak predictors. If one is weak but the other is not, the difference between the two measures can still be very large if there is high correlation between $\mathbf{X}_N$ and $\mathbf{X}_A$. Indeed, the appearance of 'one-weak and one-strong' scenarios is quite common in practice when the two predictors are highly correlated because of the well-known 'collinearity' problem among the predictors. And it is precisely in such cases that the recognition of the impact of the allowable covariates on the non-allowable ones, or *vice versa*, is of critical importance. As mentioned in Section 1, the common approach of adjusting only the allowable covariates without considering its impact on the non-allowable covariates can lead to serious misestimation of the disparity when the allowable covariates appear to be a weak predictor in the regression of Y on $\mathbf{X}_N$ and $\mathbf{X}_A$.

3.2. A discrete-distribution example

We start with a simple $2 \times 2 \times 2$ contingency table example to illustrate both the basic calculations for D_C and D_M, and their differences. We use data from the combined data set of three large epidemiological studies, namely, the NIMH Collaborative Psychiatric Epidemiology Survey (CPES): the National Latino and Asian American Study (NLAAS) [14], the National Comorbidity Study Replication (NCS-R) [15], and the National Study of American Life (NSAL) [16]. These studies focus on collecting epidemiological information on mental health and substance disorders and services utilization among the general population with special emphasis on ethnic minority groups in the NLAAS (Latinos and Asians) and NSAL (African Americans and Afro-Caribbean) with non-Latino white comparisons from the NCS-R. The studies were designed to allow integration as though they were a single, nationally representative study [17]. The combined data set is the largest epidemiological data set available for examining the patterns and correlates of

DOI: 10.1002/sim

the use of mental health services in minority populations in the United States. The sampling frames and sample selection procedures are described in detail elsewhere [18]. For illustration purposes, here we treat this combined data set as a *population* by itself; therefore, all the numbers below are regarded as population quantities (e.g. probabilities) instead of sample estimates (e.g. sample proportions).

For simplicity, we focus on a dichotomous outcome, namely, $Y=1$ means that the respondent had at least one visit to any mental health service provider (either specialist or generalist) in the past year, and $Y=0$ otherwise. The allowable covariate is also a binary variable indicating clinical need: $X_{\mathrm{A}}=1$ if there was a need, and $X_{\mathrm{A}}=0$ if there was not. The non-allowable covariate is a binary variable indicating nativity: $X_{\mathrm{N}}=1$ if the respondent is an immigrant, and $X_{\mathrm{N}}=0$ if the respondent was born in the United States.

Table I provides the data for the non-Latino white population, from which we can easily calculate the service use rate for this population. In Table I, there are two numbers in each of the cells in the 2×2 layout. The top number is the percentage of individuals who fall into the (i, j)-cell defined by the values of $(X_{\mathrm{N}}=i, X_{\mathrm{A}}=j)$, and the bottom bracketed number μ_{ij} is the percentage of people in that cell who have used services, that is, $\mu_{ij}=P(Y=1|X_{\mathrm{N}}=i, X_{\mathrm{A}}=j)$. Consequently, the overall service rate for the non-Latino white population, namely $E_1[Y]$, is obtained by multiplying the two numbers in each cell and adding them up across all cells. This leads to $E_1[Y]=14.39$ per cent. Similarly, for the Afro-Caribbean population (Table II), $E_2[Y]=6.75$ per cent, so that the observed *racial/ethnic difference* is

$$RD=E_2[Y]-E_1[Y]=6.75 \text{ per cent}-14.39 \text{ per cent}=-7.64 \text{ per cent} \tag{24}$$

This, however, is not necessarily the *disparity* in the sense of the IOM definition because it has not adjusted for the difference in clinical needs.

Comparing Tables I and II, we observe an interesting phenomenon. The percentages of people in need are greater in the Afro-Caribbean population than in the non-Latino white population when *conditional on the nativity*—55.75 per cent *versus* 41.62 per cent for the U.S. born population and 33.90 per cent *versus* 30.91 per cent for the immigrant population. The pattern, however, is *reversed* for the *marginal rates*, that is, when we combine the U.S. born and the immigrants together: 41.18 per cent for the Afro-Caribbean *versus* 41.28 per cent for the non-Latino whites. Although the difference between these two marginal rates is minimal (but there is no estimation error here as we are using the data as if they were the entire population), it is nevertheless an example of the

Table I. Non-Latino white population, where $\mu_{ij}=P_1(Y=1|X_{\mathrm{N}}=i, X_{\mathrm{A}}=j)$.

		$X_{\mathrm{A}}=$clincial need		
		No (0) (per cent)	Yes (1) (per cent)	$P_1(X_{\mathrm{A}}=1\|X_{\mathrm{N}})$ (per cent)
$X_{\mathrm{N}}=$nativity	U.S. born (0)	56.45 [$\mu_{00}=6.25$]	40.25 [$\mu_{01}=26.04$]	41.62
	Immigrant (1)	2.28 [$\mu_{10}=6.31$]	1.02 [$\mu_{11}=23.36$]	30.91
	$P_1(X_{\mathrm{A}})$	58.72	41.28	

DOI: 10.1002/sim

Table II. Afro-Caribbean population, where $\mu_{ij} = P_2(Y=1|X_N=i, X_A=j)$.

		X_A = clinical need		
		No (0) (per cent)	Yes (1) (per cent)	$P_2(X_A=1\|X_N)$ (per cent)
X_N = nativity	U.S. born (0)	14.75 [$\mu_{00}=1.19$]	18.58 [$\mu_{01}=25.61$]	55.75
	Immigrant (1)	44.07 [$\mu_{10}=1.88$]	22.60 [$\mu_{11}=4.39$]	33.90
	$P_2(X_A)$	58.82	41.18	

well-known *Simpson's paradox* [19]. The reason is the extreme imbalance of the nativity groups in the two populations: more than 95 per cent of the non-Latino whites were U.S. born, but only $\frac{1}{3}$ of the Afro-Caribbean were U.S. born.

The implication of this phenomenon for our disparity measure is clear. First, given that the difference in the marginal rates is so small, 41.18 per cent *versus* 41.28 per cent, one would expect that the *conditional disparity* that results from adjusting the Afro-Caribbean's marginal rate from 41.18 per cent to the non-Latino whites marginal rate of 41.28 per cent will have a minimal impact on the value of RD of (24). Indeed, as shown below, the *conditional disparity* in this case is $D_C = -7.62$ per cent, nearly identical to $RD = -7.64$ per cent.

Second, this adjustment in fact is in the wrong direction, because in this case the casual assumption underlying the conditional disparity, that is, the allowable covariate (clinical need) causes the non-allowable (nativity), is clearly a very implausible one. The *marginal disparity* approach is a much more sensible one, because it makes adjustment of clinical needs *within each nativity category*. Given the fact that the two nativity groups have very different levels of clinical needs, with the U.S. born having more needs, it is intuitive that we should make the adjustment after stratifying by nativity groups. Because the Afro-Caribbean population has more need in each of the nativity groups, it is also intuitive that had their needs been the same as the non-Latino whites, the observed racial/ethnic difference would be even larger. Indeed, as shown below, the marginal disparity in this case is $D_M = -8.84$ per cent. In contrast to D_C, which points to the wrong direction, D_M shows that the disparity is actually more pronounced than the unadjusted racial/ethnic difference by about $(8.84-7.64)/7.64 \approx 16$ per cent.

3.3. Disparity calculations

The calculations of D_C and D_M can be best illustrated by creating two adjusted versions of Table II, corresponding, respectively, to the two hypothetical populations as defined in (6) and (13). They are given in Tables III and IV, respectively. To construct Table III, which is for the *conditional disparity*, we need to compute the density ratio R_C of (8). From the last row of Tables I and II, respectively, we can obtain this easily as

$$R_C(0) = \frac{P_1(X_A=0)}{P_2(X_A=0)} = \frac{0.5872}{0.5882} = 0.9983, \quad R_C(1) = \frac{P_1(X_A=1)}{P_2(X_A=1)} = \frac{0.4128}{0.4118} = 1.0024$$

We can then multiply each of the three *unbracketed* proportions in the 'No (0)' column of Table II by $R_C(0)$, and multiply each of the three *unbracketed* proportions in the 'Yes (1)' column

DOI: 10.1002/sim

Table III. Adjusted Afro-Caribbean population for computing D_C.

		X_A = clinical need		
		No (0) (per cent)	Yes (1) (per cent)	$P_2^{(C)}(X_A=1\|X_N)$ (per cent)
X_N = nativity	U.S. born (0)	14.72	18.62	55.85
	Immigrant (1)	44.00	22.65	33.98
	$P_2^{(C)}(X_A)$	58.72	41.27	

Table IV. Adjusted Afro-Caribbean population for computing D_M.

		X_A = clinical need		
		No (0) (per cent)	Yes (1) (per cent)	$P_2^{(M)}(X_A=1\|X_N)$ (per cent)
X_N = nativity	U.S. born (0)	19.46	13.87	41.61
	Immigrant (1)	46.06	20.61	30.91
	$P_2^{(M)}(X_A)$	65.52	34.48	

of Table II by $R_C(1)$. This will yield the adjusted population corresponding to the conditional disparity approach, as given in Table III, where the last column $P(X_A=1|X_N)$ has also been changed using the adjusted cell probabilities. We see that Tables III and I have the same marginal distribution for X_A (rounding errors notwithstanding), as intended. The expected value of Y under this adjusted population can be easily obtained by multiplying each cell probability in Table III with the corresponding μ_{ij} from Table II and then summing them up. This leads to $E_2^{(C)}[Y]=6.77$ per cent; hence,

$$D_C = E_2^{(C)}[Y] - E_1[Y] = 6.77 \text{ per cent} - 14.39 \text{ per cent} = -7.62 \text{ per cent}$$

To calculate the marginal disparity, we need first to compute the R_M function of (15), which is determined by the rightmost columns labeled '$P(X_A=1|X_N)$' in Tables I and II. Specifically, we have

$$R_M(0;0) = \frac{P_1(X_A=0|X_N=0)}{P_2(X_A=0|X_N=0)} = \frac{1-0.4162}{1-0.5575} = 1.3193, \quad R_M(0;1) = \frac{0.4162}{0.5575} = 0.7465$$

$$R_M(1;0) = \frac{P_1(X_A=0|X_N=1)}{P_2(X_A=0|X_N=1)} = \frac{1-0.3091}{1-0.3390} = 1.0452, \quad R_M(1;1) = \frac{0.3091}{0.3390} = 0.9118$$

Table IV then is obtained by multiplying the (i, j)-cell proportion (the top unbracketed percentage) in Table II with $R_M(i; j)$ just obtained for $i, j = 0, 1$ and then compute the corresponding $P(X_A = 1|X_N)$ and $P_2^{(M)}(X_A)$ accordingly. We note that the resulting conditional distribution $P_2^{(M)}(X_A|X_N)$ is the same as that from Table I (rounding errors notwithstanding), as it should be, but the marginal

DOI: 10.1002/sim

distribution $P_2^{(M)}(X_A)$ is now markedly different from the one from the non-Latino whites, $P_1(X_A)$. This difference reflects the difference between the two approaches, because with the conditional disparity approach we have $P_2^{(C)}(X_A) = P_1(X_A)$. As we discussed previously, the seemingly natural 'equating-the-need-level' approach is actually misleading in this application because of Simpson's paradox. Equating the need level after stratifying on nativity is a much more sensible approach.

To find the expectation of Y under this adjusted Afro-Caribbean population, we multiply the four cell percentages in Table IV, respectively, by the four μ_{ij} values in Table II and then sum them up. This yields $E_2^{(M)}[Y] = 5.55$ per cent. Consequently, the marginal disparity, which in this example can be regarded as a sensible measure of disparity, is given by

$$D_M = E_2^{(M)}[Y] - E_1[Y] = 5.55 \text{ per cent} - 14.39 \text{ per cent} = -8.84 \text{ per cent}$$

3.4. A continuous-distribution example

This theoretical example establishes the mathematical fact that the difference in the conditional disparity and marginal disparity can be arbitrarily large. It also illustrates another form of Simpson's paradox, that is, even when there is no disparity in any strata defined by the non-allowable variables X_N, in the aggregated population one can still observe a disparity due to the correlation between X_N and race/ethnicity in the aggregated population and the fact that X_N is classified as non-allowable.

To see this, let us consider a simple linear regression case

$$E_k[Y|X_N, X_A] = \beta^{(k)} + \beta_N^{(k)} X_N + \beta_A^{(k)} X_A \tag{25}$$

where $k = 1$ indexes the non-Latino white population and $k = 2$ the minority population. To simplify algebra, suppose that in the natural populations (X_N, X_A) is bivariate normal, with mean $(\mu_N^{(k)}, \mu_A^{(k)})$, unit variances and correlation $\rho^{(k)}$. That is,

$$\begin{pmatrix} X_N \\ X_A \end{pmatrix}_k \sim N\left[\begin{pmatrix} \mu_N^{(k)} \\ \mu_A^{(k)} \end{pmatrix}, \begin{pmatrix} 1 & \rho^{(k)} \\ \rho^{(k)} & 1 \end{pmatrix} \right], \quad k = 1, 2 \tag{26}$$

Under this setting, for the conditional disparity, the hypothetical joint distribution $P_2^{(C)}(X_N, X_A) = P_2(X_N|X_A)P_1(X_A)$ is a bivariate normal with the following distribution:

$$\begin{pmatrix} X_N \\ X_A \end{pmatrix}_2^{(C)} \sim N\left[\begin{pmatrix} \mu_N^{(2)} + \rho^{(2)}(\mu_A^{(1)} - \mu_A^{(2)}) \\ \mu_A^{(1)} \end{pmatrix}, \begin{pmatrix} 1 & \rho^{(2)} \\ \rho^{(2)} & 1 \end{pmatrix} \right] \tag{27}$$

In contrast, under the marginal disparity approach, the hypothetical joint distribution for (X_N, X_A) is given by $P_1(X_A|X_N)P_2(X_N)$, which is also bivariate normal but with the following means and covariance matrix:

$$\begin{pmatrix} X_N \\ X_A \end{pmatrix}_2^{(M)} \sim N\left[\begin{pmatrix} \mu_N^{(2)} \\ \mu_A^{(1)} + \rho^{(1)}(\mu_N^{(2)} - \mu_N^{(1)}) \end{pmatrix}, \begin{pmatrix} 1 & \rho^{(1)} \\ \rho^{(1)} & 1 \end{pmatrix} \right] \tag{28}$$

Simple algebra then yields that the difference between the two measures is

$$\Delta D = \rho^{(2)} \beta_N^{(2)} (\mu_A^{(1)} - \mu_A^{(2)}) + \rho^{(1)} \beta_A^{(2)} (\mu_N^{(1)} - \mu_N^{(2)}) \tag{29}$$

DOI: 10.1002/sim

From (29), we have the following observations, two of which are special cases of what we have discussed in general in Section 3.1. Specifically, we see that $\Delta D = 0$ whenever one of the following three condition holds:

(a) $\rho^{(1)} = \rho^{(2)} = 0$, that is, when X_N and X_A are independent in *both* populations;
(b) $\beta_N^{(2)} = \beta_A^{(2)} = 0$, that is, when regression (25) does not depend on *either* X_N or X_A in the population of interest (not necessarily in the reference population);
(c) $\mu_N^{(1)} = \mu_N^{(2)}$ and $\mu_A^{(1)} = \mu_A^{(2)}$, that is, when the two populations have the same marginal distributions for *both* X_N and X_A.

Of course ΔD can be zero by many other (incidental) combinations of the parameter values, but the above three are most useful for theoretical insights. Note in particular that conditions (a) and (b) are applicable in general, but condition (c) works only when the regression of Y is linear in both X_N and X_A. We emphasize that as the parameters in (29) have no restrictions other than $|\rho^{(k)}| \leqslant 1$, ΔD can be arbitrarily large, including approaching infinity.

We also remark a special case of interest, that is, when $E_k[Y|X_N, X_A]$ of (25) is free of both k (e.g. race/ethnicity index) and X_N (i.e. $\beta_N^{(k)} = 0$). In such cases, there is no racial/ethnic disparity under the conditional disparity model, as X_A is being adjusted to have the same distribution for both racial/ethnic groups and (11) does not involve X_N. Under the marginal disparity model, however, the matter is more complicated. Although X_N does not impact Y directly, it impacts X_A when it is correlated with X_A. Consequently, the difference in the marginal distributions of X_N in the two racial/ethnic groups will result in differences in the marginal distribution of X_A even when, or rather especially when, the conditional distribution $P_k^{(N)}(X_A|X_N)$ is adjusted to be invariant to the race/ethnicity index k. It then follows that there will be a racial/ethnic disparity due to the indirect impact of X_N on Y via X_A. Indeed, it is easy to verify for the current example that the marginal disparity is given by

$$D_M = \rho^{(1)} \beta_A^{(2)} (\mu_N^{(2)} - \mu_N^{(1)}) \tag{30}$$

This is zero only when (i) $\rho^{(1)} = 0$ and hence X_A and X_N are independent in the reference population; therefore, X_N cannot impact X_A in the hypothetical population, (ii) $\beta_A^{(2)} = 0$ and hence the impact of X_N on X_A does not translate into any impact on Y in the hypothetical population, or (iii) $\mu_N^{(2)} = \mu_N^{(1)}$ and hence the distribution of X_N is actually invariant to race/ethnicity.

Perhaps most important here is to note Simpson's paradox again. Although in the aggregated population there is a marginal disparity for the case above, clearly there is no disparity in any subpopulation defined by a particular value of X_N, that is, when we condition on X_N, because the conditional distribution $P_2(X_A|X_N)$ has been adjusted to be the same as $P_1(X_A|X_N)$. This of course is not paradoxical, just as Simpson's paradox is not a real paradox in the mathematical sense. Once we classify X_N as a non-allowable variable, then logically we have to accept any difference caused by it as a part of the overall disparity, regardless of whether the difference comes from its direct impact or indirect impact on the outcome Y. Of course, one may argue whether the indirect part really should be viewed as disparity, which is not an easy issue to address as then one is implying that X_N is both a non-allowable variable (for the direct impact) and allowable variable (for the indirect impact via X_N). We shall pursue this complex issue in subsequent work.

DOI: 10.1002/sim

4. FUTURE WORK

The IOM definition of disparities takes an indirect approach of elimination and defines health care disparity as the difference in health care that is *not due to* allowable covariates. While this approach is appropriate for capturing disparity in its entirety irrespective of source attribution, it leaves open the question of plausible causes for the disparity, and what can be done to eliminate or reduce the disparity.

An alternative direct constructive approach is to define health care disparity attributable to specific non-allowable covariates as the difference in health care that is *due to* these covariates. This approach is used widely in the literature on attributable risk, such as Rubin [20], which also emphasized the importance of making explicit assumptions such as what we termed the preservation assumptions. This alternative approach can be implemented using the similar statistical frameworks proposed above, but with the role of allowable and non-allowable covariates switched. This approach does not capture disparity in its entirety, because it captures only disparity attributable to the specific non-allowable covariates and may miss the disparity attributable to other non-allowable covariates, including those that may not have been observed. However, this approach may have more direct policy implications, providing guidance on the potential to reduce or even eliminate health care disparities through specific policy implementations regarding the specific non-allowable covariates.

In practice, we believe that both versions of the disparity are important. The elimination approach is useful for estimating the magnitude of the overall disparity, whereas the constructive approach is a tool for estimating how much disparity can be eliminated through specific policy interventions. A comparison between the two is also important in revealing how much of the overall disparity the policy intervention can eliminate. If a large portion remains, a new policy intervention needs to be identified. We plan to explore these issues in subsequent work, especially in the context of longitudinal data.

Another issue that we plan to investigate is the issue of variables that are not included in the model for predicting the outcome Y but may actually be important. Traditionally there is not much one can do about those variables other than trying one's best to include as many variables as one can find and afford to measure. For the conditional disparity framework as we outlined, one may have noted that the conditional disparity as defined by (11) does not involve the non-allowable variables. This provides an opportunity to realize the implicit assumption carried in the IOM definition, that is, the non-allowable category is the 'catch all' category that includes all covariates that have not been named explicitly in the allowable category. Of course, without strong assumptions, nothing can be done for variables that are not even identified. Recall the fundamental assumption underlying our conditional disparity model, which is that the allowable variables, which clearly need to be identified and measured, are causes for non-allowable variables. Therefore, if in specific applications where such an assumption can be viewed as reasonable, even when the non-allowable variables form the 'catch all' category, then the conditional disparity measure enjoys the property of being more general than we discussed in this paper.

However, the 'catch-all' formulation of the non-allowable variables would not produce anything meaningful under the marginal disparity model, because neither can we stratify on variables that are not measured nor should it be as logically nothing can be done when the causes are not even identified. All these issues remind us again of the fundamental importance of explicitly formulating, identifying, and stating causal assumptions underlying any disparity measure.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

We thank J. Gastwirth, D. Rubin, X. Xie, and A. Zaslavsky for their helpful exchanges.

REFERENCES

1. Institute of Medicine. *Unequal Treatment*: *Confronting Racial and Ethnic Disparities in Health Care*. National Academy Press: Washington, DC, 2002.
2. Asch DA, Armstrong K. Aggregating and partitioning populations in health care and partitioning populations in health care disparities research: differences in perspective. *Journal of Clinical Oncology* 2007; **25**(15):2117–2121.
3. Cook B. Effect of medicaid managed care on racial disparities in health care access. *Health Services Research* 2007; **42**:124–145.
4. McGuire TG, Alegria M, Cook BL, Wells KB, Zaslavsky AM. Implementing the Institute of Medicine definition of disparities: an application to mental health care. *Health Services Research* 2006; **41**:1979–2005.
5. Rao RS, Graubard BI, Breen N, Gastwirth JL. Understanding the factors underlying disparities in cancer screening rates using the Peters–Belson approach. *Medical Care* 2004; **42**(8):789–800.
6. Fiscella K, Franks P, Doescher MP, Saver BG. Disparities in health care by race, ethnicity, and language among the insured: findings from a national sample. *Medical Care* 2002; **40**(1):52–59.
7. Gastwirth JL. A clarification of some statistical issues in Watson V. Fort Worth Bank and Trust. *Jurimetrics Journal* 1989; **29**:267–284.
8. Gastwirth JL, Greenhouse SW. Biostatistical concepts and methods in the legal setting. *Statistics in Medicine* 1995; **14**:1641–1653.
9. Nayak TK, Gastwirth JL. Statistical measures of economic discrimination useful in evaluating fairness. *Proceedings of Biopharmaceutical Section*. American Statistical Association: Alexandria, VA, 1995; 87–94.
10. Gelman A, Meng X-L (eds). *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives*. Wiley: U.K., 2004.
11. Holland PW, Wang YJ. Dependence function for continuous bivariate densities. *Communications in Statistics Part A—Theory and Methods* 1987; **16**:863–876.
12. Wang YJ. Construction of continuous bivariate density functions. *Statistica Sinica* 1993; **3**:173–187.
13. Molenberghs G, Lesaffre E. Non-linear integral equations to approximate bivariate densities with given marginals and dependence functions. *Statistica Sinica* 1997; **7**:713–738.
14. Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng X-L, Vega W, Zane N, Vila D, Woo M, Vera M, Guarnaccia P, Aguilar-Gaxiola S, Sue S, Escobar J, Lin K, Gong F. Considering context, place and culture: the National Latino and Asian American study. *International Journal of Methods in Psychiatric Research* 2004; **13**(4):208–220.
15. Kessler R, Merikangas K. The National Comorbidity Survey Replication (NCS-R). *International Journal of Methods in Psychiatric Research* 2004; **13**(2):60–68.
16. Jackson J, Torres M, Caldwell C, Neighbors H, Nesse R, Taylor RJ, Treirweiler S, Williams DR. The National Survey of American Life: a study of racial, ethnic and cultural influences on mental disorders and mental health. *International Journal of Methods in Psychiatric Research* 2004; **13**(4):196–207.
17. Heeringa S, Berglund P. *National Institutes of Mental Health* (*NIMH*) *Data Set*, *Collaborative Psychiatric Epidemiology Survey Program* (*CPES*): *Integrated Weights and Sampling and Sampling Error Codes for Design-based Analysis*. http://www.icpsr.umich.edu/cocoon/cpes/using.xml?section=Weighting, 2007.
18. Heeringa S, Wagner J, Torres M, Duan N, Adams T, Berglund P. Sample designs and sampling and sampling methods for the Collaborative Psychiatric Epidemiology Studies (CPES). *International Journal of Methods in Psychiatric Research* 2004; **13**(4):221–240.
19. Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, *Series B* 1951; **13**:238–241.
20. Rubin DB. Estimating the causal effects of smoking. *Statistics in Medicine* 2001; **20**(9–10):1395–1414.

DOI: 10.1002/sim